Correcting Misperceptions about Support for Social Distancing to Combat COVID-19†

James Allen IV[1,2,3], Arlete Mahumane[4], James Riddell IV[5], Tanya Rosenblat[1,6], Dean Yang[1,2,3], and Hang Yu[7,8]

[1]Department of Economics, University of Michigan
[2]Ford School of Public Policy, University of Michigan
[3]Population Studies Center, University of Michigan
[4]Beira Operational Research Center, National Institute of Health, Mozambique
[5]Division of Infectious Diseases, University of Michigan Medical School
[6]School of Information, University of Michigan
[7]National School of Development, Peking University
[8]Institute of South-South Cooperation and Development, Peking University

Version: November 23, 2021

Abstract

Can informing people of high community support for social distancing encourage them to do more of it? We randomly assigned a treatment correcting individuals' underestimates of community support for social distancing. In theory, informing people that more neighbors support social distancing than expected encourages *free-riding* and *lowers* the perceived benefits from social distancing. At the same time, the treatment induces people to revise their beliefs about the infectiousness of COVID-19 upwards; this *perceived infectiousness* effect as well as the *norm adherence* effect *increase* the perceived benefits from social distancing. We estimate impacts on social distancing, measured using a combination of self-reports and reports of others. While experts surveyed in advance expected the treatment to increase social distancing, we find that its average effect is close to zero and significantly lower than expert predictions. However, the treatment's effect is heterogeneous, as predicted by theory: it decreases social distancing where current COVID-19 cases are low (where free-riding dominates), but increases it where cases are high (where the perceived-infectiousness effect dominates). These findings highlight that correcting misperceptions may have heterogeneous effects depending on disease prevalence.

JEL Classification: I12, D91, O12

Keywords: COVID-19, Social Distancing, Health Behavior, Health Policy, Behavioral Economics

†Contacts: alleniv@umich.edu; deanyang@umich.edu. Faustino Lessitala provided top-notch leadership and field management. Patricia Freitag, Ryan McWay, and Maggie Barnard provided excellent research assistance. Julie Esch, Laura Kaminski, and Lauren Tingwall's grant management was world-class. We appreciate feedback from participants in Michigan's Health, History, Development, and Demography (H2D2) Seminar and the Conference on Experimental Insights from Behavioral Economics on COVID-19 (JHU-LSE). This work is supported by the Abdul Latif Jameel Poverty Action Lab (J-PAL) Innovation in Government Initiative through a grant from The Effective Altruism Global Health and Development Fund (grant number IGI-1366), the UK Foreign, Commonwealth & Development Office awarded through Innovations for Poverty Action (IPA) Peace & Recovery Program (grant number MIT0019-X9), the Michigan Institute for Teaching and Research in Economics (MITRE) Ulmer Fund (grant number G024289), and the National Institute on Aging of the National Institutes of Health (award number T32AG000221). Our protocols were reviewed and approved by Institutional Review Boards (IRBs) at the University of Michigan (Health Sciences and Social and Behavioral Sciences IRB, approval number HUM00113011) and the Mozambique Ministry of Health National Committee on Bioethics for Health (CNBS reference number 302/CNBS/20). The study was submitted to the AEA RCT Registry on May 26, 2020, registration ID number AEARCTR-0005862: 10.1257/rct.5862-3.0. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the aforementioned institutions.

1 Introduction

Attitudes toward social distancing have changed rapidly during the pandemic (Janzwood, 2020). During such rapid change, people often underestimate support for social distancing in their communities. In our Mozambican sample, 98% thought that people should be social distancing, but estimated that only 69% of others in the community felt similarly. This gap motivates a public health policy: simply inform people of high rates of community support for social distancing. What impact would such messaging have on social distancing behavior?

In theory, the impact of such a "misperceptions correction" intervention on social distancing is ambiguous: on the one hand, informing people that more neighbors support social distancing than expected encourages *free-riding* and *lowers* the perceived benefits from social distancing. On the other hand, people should revise their belief about the seriousness of COVID-19 upwards in order to rationalize the observed number of infections in their neighborhood despite the higher than expected social distancing support. This *perceived infectiousness* effect *increases* the perceived benefits from social distancing and dominates free-riding in communities with high levels of infections.[1] Finally, the *norm adherence* effect should induce people to follow whatever local social norm is set by their neighbors - in our case this effect should always increase social distancing.

We implemented a randomized controlled trial testing the impact of informing people about high local support for social distancing. The treatment either updated beliefs upwards or confirmed beliefs about high rates of support for social distancing. Abiding by COVID-19 protocols, we conducted all treatments and surveys by phone among 2,117 Mozambican households.

Our outcome variable is the extent to which a household engages in social distancing. Measuring this behavior is challenging due to experimenter demand effects.[2] Yet most prior studies ask for self-reports about general social distancing compliance. When we do so, 95% claim to observe government social distancing recommendations. We therefore construct a novel measure of social distancing. First, we ask respondents to self-report several social distancing actions. Second, we ask *others* in the community to report on the respondent's social distancing. We are aware of no prior study that makes use of other-reports on a respondent's social distancing behavior. Incorporating self-reported actions and others' reports drops social distancing to a more discerning 8% (see Figure 1 and Section 3.2). Improved measurement leads the social distancing rate to fall by an order of magnitude.

The average effect of the misperceptions correction treatment in the full sample is small

[1] Our model is related to the literature on decision-making under misspecified subjective models (Spiegler, 2020). Agents hold incorrect assumptions on one model parameter (e.g., share of population social distancing), leading them to incorrect conclusions about other parameters (e.g., disease infectiousness).

[2] Jakubowski et al. (2021) find that self-reported mask wearing is overstated relative to measures based on observations of others.

and not statistically significantly different from zero. However, as theory predicts, there is substantial treatment effect heterogeneity: the treatment effect is statistically significantly more positive when local COVID-19 cases (per 100,000 population) are higher. In districts with few cases, the treatment effect is negative. In the district with the most COVID-19 cases, the treatment increases social distancing by 9.3 percentage points (statistically significant at the 5% level), a 75% increase over that district's control-group mean.

This pattern is consistent with the theoretical prediction that as infection rates rise, the perceived-infectiousness effect should increasingly dominate the free-riding effect of the misperceptions correction treatment, leading the treatment effect to become more positive. We also test a further implication of the model: expectations of future infection rates should show similar treatment effect heterogeneity. Empirical analyses confirm this prediction, providing additional support for the theoretical model.

This paper contributes to understanding the impact of providing information about others' beliefs and attitudes (Benabou and Tirole, 2011; Bicchieri and Dimant, 2019). In health settings, Yu (2020) and Yang et al. (2021) find (in an overlapping Mozambican sample) that correcting overestimates of stigmatizing attitudes promoted HIV testing, though Banerjee et al. (2019) find that informing Nigerian young adults of peers' attitudes on healthy sexual relationships did not change respondents' own attitude.[3] Regarding social distancing, Martinez et al. (2021) show that respondents are influenced by others' social distancing actions in hypothetical vignettes; however, no prior study has tested the impact of providing information on community support of social distancing on respondent behavior.

Our emphasis on interactions between free-riding and perceived-infectiousness effects is novel, but each effect has been studied separately. Free-riding has been studied in the context of vaccination decisions (Hershey et al., 1994; Lau et al., 2019) and social distancing (Cato et al., 2020). Perceived COVID-19 infection risk (e.g., due to vaccine anticipation, Andersson et al. (2021)) has been shown to lower social distancing intentions.

2 Theory

Our model focuses on the interaction between the free-riding and perceived infectiousness effects for communities with low and high overall infection rates. We view norm-adherence as a uniform effect that should always increase social distancing.

We consider a community where people have random pairwise meetings. People believe that a share x of the population supports social distancing and that the probability of becoming infected from unprotected meetings is α. People treat x as given, but infer the infectiousness α from the current infection rate R in the community which they can observe (we describe this inference below). The true infectiousness of the disease is $\hat{\alpha}$.

[3] Misperception-correcting interventions have also been shown to change energy consumption (Schultz et al., 2007) and female labor force participation (Bursztyn et al., 2018).

Importantly, people in the community have *miscalibrated beliefs*: the true share of the population supporting social distancing is $\hat{x}$ (we are interested in the case $\hat{x} > x$). People infer the true infectiousness $\hat{\alpha}$ of the disease only if they are correctly calibrated ($x = \hat{x}$).

Individual Effort. A supporter engages in preventative effort e and assumes that other supporters choose effort e^* (in equilibrium we have $e = e^*$). Non-supporters choose effort $e = 0$.

When someone supporting social distancing meets another person, she escapes exposure with probability:

$$\begin{aligned} A(e, e_{other}) &= \sqrt{e + e_{other}} \\ &= \begin{cases} \sqrt{e + e^*} & \text{if other person is supporter} \\ \sqrt{e} & \text{if other person is non-supporter} \end{cases} \end{aligned} \tag{1}$$

Hence, the marginal benefit of effort decreases both with own effort e as well as the other person's effort e^*.[4]

The expected probability of escaping exposure is therefore:

$$\overline{A}(e, e^*) = (1 - x)\sqrt{e} + x\sqrt{e + e^*} \tag{2}$$

An agent becomes exposed with probability $1 - \overline{A}(e, e_{other})$. If exposed she gets infected with probability α and suffers disutility $-C$ from infection.[5] If she is not exposed then she does not get infected. Her baseline utility from no infection equals $\overline{U}$. The cost of preventative effort is e. Hence, her total utility equals:

$$\overline{U} - \alpha(1 - \overline{A}(e, e_{other}))C - e \tag{3}$$

The agent chooses e to maximize her utility, giving us the following first-order condition:

$$\frac{\alpha C}{2\sqrt{e}} \left[1 - x(1 - \frac{1}{\sqrt{1 + \frac{e^*}{e}}}) \right] = 1 \tag{4}$$

In equilibrium it has to be the case that the population effort e^* equals e. Hence, we can

[4] We assume the other person's effort is unobservable. This is consistent with our finding that respondents underestimate the extent of social distancing.

[5] For simplicity, we assume that infectiousness does not vary with the agent's type (supporter or non-supporter). Otherwise, we would need to keep track of two levels of infectiousness. The qualitative results would not change.

characterize equilibrium effort as:

$$e = \left(\frac{\alpha C}{2}\left[1 - x(1 - \frac{1}{\sqrt{2}})\right]\right)^2 \tag{5}$$

This demonstrates the *free-riding effect*: increasing the share x of supporters *decreases* effort because the marginal utility from own effort decreases. Also, effort increases if the disease is more infectious (higher α) and if illness is costlier (higher C).

Infection Rate. People observe the current infection rate in the community. Infections come from two sources: non-supporters become sick at rate $\alpha\,(1 - x\sqrt{e})$ while supporters become sick at rate $\alpha(1 - \overline{A}(e, e))$. Hence, people in the community assume that the current infection rate is generated by the following process:

$$\begin{aligned} R &= \alpha \left[\underbrace{(1-x)(1-x\sqrt{e})}_{\text{non-supporters}} + \underbrace{x\left(1 - \sqrt{e}(1 + (\sqrt{2}-1)x)\right)}_{\text{supporters}}\right] \\ &= \alpha \left[1 - \sqrt{e}\,\underbrace{2x\left(1 - x\left(1 - \frac{1}{\sqrt{2}}\right)\right)}_{=G(x)}\right] \end{aligned} \tag{6}$$

However, the true process determining current infections is actually:

$$R = \hat{\alpha}\left[1 - \sqrt{e}G(\hat{x})\right] \tag{7}$$

In other words, the true infection process is driven by the same social distancing effort of supporters but different infectiousness $\hat{\alpha}$ and different $\hat{x}$.

2.1 Basic Equilibrium

Supporters initially assume that the disease has low infectiousness and they adjust their estimate of α upwards until the current infection rate R stabilizes.

Proposition 1 *In equilibrium, effort level e, the current infection rate R, and the assumed infectiousness α satisfy Equations 5, 6 and 7. Moreover, $\hat{\alpha} > \alpha$ if $\hat{x} > x$.*

In equilibrium, both the assumed infection process (Equation 6) and the real infection rate (Equation 7) must produce observed infection rate R. For the second part, note that $G(x)$ is increasing in $x \in [0, 1]$: hence, $\hat{x} > x$ implies $\hat{\alpha} > \alpha$ to generate the same infection rate R.

2.2 Treatment Effect

We now consider the effect of our treatment informing people that the population share supporting social distancing is really $\hat{x} > x$.

Proposition 1 implies that if supporters are informed that the true population share supporting social distancing is $\hat{x} > x$, they must infer higher disease infectiousness than they initially assumed (because their estimated disease infectiousness immediately jumps from α to true $\hat{\alpha}$). This is the *perceived-infectiousness effect.*

Supporters of social distancing will adjust their effort level to a new level $\hat{e}$, but there are two countervailing effects:

1. Holding assumed infectiousness α constant, the free-riding effect *decreases* effort.

2. The perceived-infectiousness effect *increases* effort, because the agent now believes the disease is more infectious than initially thought (perceived α increases), increasing the gain from social distancing.

Intuitively, the perceived-infectiousness effect varies monotonically with R: when infections are low, supporters' effort is low, and both supporters and non-supporters get infected at similar rates. Hence, agents revise the estimate of infectiousness α only slightly upwards in response to the treatment. On the other hand, when infections are high, supporters' effort is high and the upward revision will be larger.

The following theorem makes this intuition precise. Instead of doing comparative statics on R (which is determined in equilibrium) we state the comparative statics results in terms of the infectiousness $\hat{\alpha}$ (for given x and $\hat{x}$). Note that R increases with $\hat{\alpha}$.

Theorem 1 *Assume an agent is informed that a share $\tilde{x} > x$ of the population supports social distancing. Then there is a threshold $\hat{\alpha}^*$ such that for any $\hat{\alpha} < \hat{\alpha}^*$ the free-riding effect dominates and equilibrium effort decreases, and for $\hat{\alpha} > \hat{\alpha}^*$ the perceived-infectiousness effect dominates and the equilibrium effort increases.*

See Appendix A for the proof.

The interplay between free-riding and perceived-infectiousness effects also yields analogous predictions about a central belief about COVID-19: the future infection rate. In the endline survey, we ask respondents to estimate this. The expected future rate differs from the current infection rate R, because this study occurs at a point when infection rates are clearly evolving. The misperceptions correction treatment changes respondent beliefs about social distancing support and about infectiousness, and therefore should change expected future infection rates.

Recall that non-supporters are always infected with higher probability than supporters. The higher the infectiousness parameter $\hat{\alpha}$, the higher should be future infection rates for both groups.

When $\hat{\alpha}$ is currently small, the perceived-infectiousness effect is small. Simultaneously, the treatment corrects beliefs about the share of social-distancing supporters upwards, which should *reduce* estimates of future infection rates because supporters have lower infection rates. Thus, the expected future infection rate *decreases* when $\hat{\alpha}$ is currently small.

In contrast, when $\hat{\alpha}$ is currently large, the treatment leads to a large increase in perceived infectiousness, implying that the disease will infect higher shares of both supporters and non-supporters. This will tend to *increase* expected future infection rates.

To summarize, the misperceptions correction treatment effect on the expected future infection rate should show heterogeneity similar to that described in Theorem 1. The treatment effect on the expected future infection rate is strictly negative if the current local infection rate (R) (which moves monotonically with $\hat{\alpha}$) is small enough. The treatment effect on the expected future infection rate increases with the current infection rate, and can become positive if current infection rates are sufficiently high.

In our empirical analyses, we test these predictions regarding heterogeneity in the misperceptions correction treatment effect.

3 Sample and Data

3.1 Data

We implemented surveys by phone in July-November 2020. Respondents were from households with phones in the sample of a prior study (Yang et al., 2021).[6] We surveyed one adult per household. Appendix B provides details on the COVID-19 context, study communities and the study timeline.

Between a pre-baseline survey and baseline survey, we randomly assigned households to treatments and registered a pre-analysis plan (PAP). The baseline survey was immediately followed by over-the-phone treatment implementation. There was a minimum of 3.0 weeks and average of 6.3 weeks between baseline and endline surveys for all respondents. Baseline and endline surveys occurred when COVID-19 cases were rising rapidly.

The endline sample size is 2,117 respondents, following a sample size of 2,226 at baseline. The retention rate between baseline and endline is 95.1% overall, at least 94.4% in each of the seven districts surveyed, and balanced across treatment conditions. Table 1 presents summary statistics. 99% of respondents support social distancing, but respondents underestimate the community share expressing such support, on average estimating 69% in a pre-baseline survey and 80% at baseline.

[6] AEA RCT Registry for Yang et al. (2021): https://doi.org/10.1257/rct.3990-5.1

3.2 Primary Outcome

The primary outcome is an indicator that the respondent practiced social distancing, as pre-specified in our PAP. It is constructed from self-reports of social distancing as well as others' reports of the respondent's social distancing. The outcome is equal to one if the respondent is practicing social distancing according to both self-reports and other-reports, and zero otherwise.

Respondents are social distancing according to their self-report if both of the following are true: 1) they answer "yes" to "In the past 14 days, have you observed the government's recommendations on social distancing?", and 2) they report doing at least seven out of eight "social distancing actions" in the past seven days (higher than the sample median number, six).[7] A list of the social distancing actions and their corresponding summary statistics are presented in Appendix F.

To collect others' reports on a respondent's social distancing, study participants were asked about their social interactions with ten other community study participants. These ten others were identified from social network data and geographic proximity. Additionally, community leaders were also asked about social interactions with all community study participants.[8] Other-reports were collected at baseline and endline.

In collecting other-reports, we asked others whether they had seen anyone from the respondent household in the last 14 days. If so, we then asked: 1) Did he/she come closer than 1.5 meters to you or others not of his/her household at any point in the last 14 days?; 2) Did he/she shake hands, try to shake hands, or touch you or others not of his/her household in the last 14 days?; and 3) In general, did he/she appear to be observing the government's recommendations on social distancing (avoid large gatherings and keep at least 1.5 meters distance from people not of his/her household)? Respondents are considered to be social distancing according to others if all others responded "no", "no", and "yes" (respectively) to these three questions, reported having not seen the respondent in the past 14 days, or reported not knowing the respondent.[9]

Figure 1 displays how these questions lead to the social distancing outcome. 95% of respondents say "yes" to the self-report on general social distancing. When considering self-reports of doing at least seven out of eight social distancing actions, the social distancing rate falls to 36%. Finally, incorporating others' reports reduces the rate further to 8%. Incorporating additional information into the social distancing measure – using self-reports of specific social distancing behaviors as well as other-reports – leads to substantially lower

[7] While this threshold was pre-specified, regression results (in Appendix K) are robust to alternate definitions of this component, such as a threshold of six, or dropping social distancing actions 4 and 6 for which respondents might misinterpret and answer "no" if not showing symptoms.

[8] Each respondent household was known at baseline by 0.98 community leaders and 3.21 neighboring survey respondents on average.

[9] At baseline, 90.55% of respondent households were known by some other respondent or community leader.

social distancing rates.

4 Research Design

4.1 Treatments

We implemented a randomized controlled trial estimating impacts on social distancing of two treatments: 1) misperceptions correction, and 2) leader endorsement. Before the baseline survey, we randomly assigned 30% of households completing the pre-baseline survey each to one of two treatments and the remaining 40% to a control group. Sample sizes by treatment condition were as follows: misperceptions correction (N=628, 29.7% of sample), leader endorsement (N=637, 30.1%), and control group (N=852, 40.3%). The leader endorsement treatment, which is found to be ineffective, is described in Appendix C.

For the misperceptions correction treatment, the pre-baseline survey asked all respondents whether they themselves support social distancing, to calculate the share in the community supporting social distancing. We then asked individuals at baseline to estimate that share (reported as an integer out of 10). Immediately after completing the baseline survey, treated individuals underestimating the share were told the true share supporting social distancing, rounded to an integer out of 10. Treated individuals correctly estimating the share were told that they were correct. In practice, 92.4% of treated respondents received this treatment, 53.2% of whom underestimated community support for social distancing and 46.8% of whom correctly estimated it. The small minority overestimating the share were not provided additional information.[10] Appendix C contains additional details and treatment scripts.

Attrition between baseline and endline is low (4.9%). In Appendix E, we show that attrition and key baseline variables are balanced across treatment conditions. Further, at endline, 97.9% recall receiving the baseline survey and, of those, 99.4% report trusting the COVID-19 information we provided.[11]

4.2 Regressions

A pre-specified ordinary-least-squares regression equation provide treatment effect estimates:[12]

$$Y_{ijd} = \beta_0 + \beta_1 T1_{ijd} + \beta_2 T2_{ijd} + \eta B_{ijd} + \delta_{ijd}^{others} + \delta_{ijd}^{leaders} + \gamma_{jd} + \varepsilon_{ijd} \tag{8}$$

[10]While respondents were not incentivized to truthfully guess community support (for scalability), true beliefs can still be updated for all except those who overestimated true community support with an upward biased guess; however, the latter case should only attenuate our treatment effect and not bias it upward.

[11]Trust may have arisen from multiple in-person household surveys since 2017 (see Yang et al. (2021))

[12]Appendix G shows that all conclusions are robust to logit and probit specifications.

where Y_{ijd} is the social distancing indicator for respondent i in community j and district d; $T1_{ijd}$ and $T2_{ijd}$ are indicator variables for the misperceptions correction and leader endorsement treatment groups, respectively; B_{ijd} is the baseline value of the dependent variable; δ_{ijd}^{others} is a vector of dummy variables for the number of other respondents who report knowing the respondent's household from 0 to 8; $\delta_{ijd}^{leaders}$ is a vector of indicators for the number of community leaders who report knowing the respondent's household from 0 to 4;[13] γ_{jd} are community fixed effects; and ε_{ijd} is a mean-zero error term. We report robust standard errors.

Coefficients β_1 and β_2 represent the intent-to-treat impacts of the misperceptions correction and leader endorsement treatments (respectively) on social distancing.

We modify Equation 8 to estimate heterogeneity in treatment effects with respect to local COVID-19 case loads:

$$\begin{aligned} Y_{ijd} = \beta_0 + \beta_1 T1_{ijd} + \beta_2 T2_{ijd} + \beta_3(T1_{ijd} * Covid_d) + \beta_4(T2_{ijd} * Covid_d) \\ + \eta B_{ijd} + \delta_{ijd}^{others} + \delta_{ijd}^{leaders} + \gamma_{jd} + \varepsilon_{ijd} \end{aligned} \tag{9}$$

Equation 9 adds interactions between treatment indicators and the cumulative number of district-level COVID-19 cases per 100,000 population at the start of the endline survey.[14] Coefficients β_1 and β_2 in Equation 9 now represent the impacts of the treatments in districts where COVID-19 cases are zero (slightly out of sample); β_3 and β_4 represent the change in the respective treatment effect for a one-unit increase in district-level COVID-19 cases per 100,000 population.

We show robustness to clustering standard errors in Appendix L. Clustering by 76 communities or 7 districts has minimal impact on standard errors, and does not affect whether any coefficients are statistically significant at conventional levels.

4.3 Hypotheses

We pre-specified the hypothesis that each treatment (β_1 and β_2 in Equation 8) would have positive effects. Subject-matter experts (surveyed without knowing results) concurred with this expectation.[15] The mean expert predictions were that the misperceptions correction and leader endorsement treatments would increase social distancing by 5.23 and 5.56 percentage points, respectively.

We also test the hypotheses that the impact of the misperceptions correction treatment on social distancing and on the expected future infection rate will be more positive the

[13] As pre-specified, we cap δ_{ijd}^{others} at the first integer that covers over 90% of the sample, and $\delta_{ijd}^{leaders}$ at the maximum number of leaders found in any community.

[14] The main effect of $Covid_d$ is absorbed by γ_{jd}.

[15] 71 individuals provided predictions at https://socialscienceprediction.org/ (survey closing date January 2, 2021).

higher the current COVID-19 infection rate (β_3 in Equation 9 will be positive). We did not pre-specify these hypotheses, but advance them on the basis of our theoretical model.

5 Results

5.1 Average Treatment Effects

In Table 2 Column (1), we present regression estimates for our primary outcome.[16] Both treatment coefficients are small in magnitude and neither is statistically significantly different from zero. These findings diverge from expert predictions of treatment effects. We strongly reject the null that our T1 and T2 treatment effect estimates are equal to the positive mean expert predictions (p-value$<$0.001 in each case).

However, we find the misperceptions correction has a positive effect on measures of perceived community support for social distancing. Analyses presented in Appendix H (not pre-specified) shows that the treatment effect is concentrated on the lower end of the distribution, having a significant positive effect on a respondent perceiving that at least 50% of households in their community support social distancing.

5.2 Treatment Effect Heterogeneity

In Table 2 Column (2), we present regression estimates of treatment effect heterogeneity (Equation 9) with respect to the local infection rate (COVID-19 cases per 100,000 population in the respondent's district).

The misperceptions correction treatment effect is heterogeneous with respect to local COVID-19 cases. The coefficient on the interaction term with $T1_{ijd}$ is positive and statistically significant at the 1% level. The coefficient on the $T1_{ijd}$ main effect is the predicted effect of misperceptions correction in a district with zero cases (slightly out of sample), and suggests that the misperceptions correction would reduce social distancing by 3.4 percentage points in such a location (statistically significant at the 5% level).

Figure 2 displays this treatment effect heterogeneity. We plot district-specific treatment effects (estimating Equation 8 separately in each of seven districts) on the y-axis (with 95% confidence intervals) against district case counts on the x-axis. In the six districts with the lowest case counts, coefficients are negative. By contrast, in Chimoio, the district with the most cases (39.08/100,000) that also accounts for one-quarter of the sample, we estimate a large positive effect: 9.3 percentage points (a 75% increase over that district's control group, statistically significant at the 5% level).

This heterogeneous treatment effect finding does not depend on inclusion of Chimoio in the sample. When estimating Equations 8 and 9 for the sample excluding Chimoio

[16]The complete set pre-specified analyses are presented in Appendix D.

(Appendix I), results are very similar.

By contrast, the leader endorsement treatment effect is not heterogeneous with respect to local case loads. The coefficient on the corresponding interaction term in Column (1) is small in magnitude and not statistically significantly different from zero.

The interplay between the free-riding and perceived-infectiousness effects is the distinctive feature of our theoretical model. When the perceived-infectiousness effect is large enough, it overcomes the countervailing free-riding effect, and the misperceptions correction treatment leads to more social distancing. An additional implication of the theory is that the treatment should have similar heterogeneous effects on the expected future infection rate.

We conduct this additional test of the theory, examining treatment effects on the expected future infection rate.[17] In Columns (3) and (4) of Table 2, the outcome is the share of the community the respondent thinks will get sick from COVID-19 (responses were integers out of 10; we divide by 10 to yield a 0-1 scale). In Column (3), we estimate average treatment effects. Each coefficient is small in magnitude and not statistically significantly different from zero.

In Column (4), we estimate heterogeneity in treatment effects with respect to local cases, and find the same pattern as in Column (2). The misperceptions correction decreases the expected future infection rate in districts with no cases, and this impact becomes more positive as current cases rise (the $T1_{ijd}$ main effect and interaction term coefficients are both statistically significant at the 5% level).

These treatment effect heterogeneity findings (Columns 2 and 4, Table 2) jointly support the theoretical model. When current infection rates are low, the misperceptions correction treatment does not change perceived infectiousness much, but leads to realizations that social distancing support is higher than previously thought. People therefore reduce estimates of the future infection rate, and also reduce their own social distancing (choosing to free-ride). By contrast, when current infection rates are high, the treatment causes larger increases in perceived infectiousness. Notwithstanding an increase in the share of social distancing supporters, people increase their estimate of the future infection rate, and increase their social distancing.

6 Conclusion

Support for social distancing increased rapidly during the COVID-19 pandemic. If people are unaware of the extent to which others' beliefs on social distancing have changed, would

[17] The question is "For every 10 people in your community, how many do you think would get sick from coronavirus?" Sample sizes in these regressions are smaller. We implemented this question midway through the endline survey, after finding preliminary evidence suggesting the need to explore mechanisms behind treatment effect heterogeneity.

revealing true high rates of such support lead to more social distancing? In theory, the impact of providing such information is ambiguous: it could reduce social distancing if free-riding effects dominate, but could have a positive effect on social distancing if perceived-infectiousness effects dominate. Perceived-infectiousness effects are more likely to dominate when the current local infection rate is higher.

We implemented a randomized controlled trial testing the impact of a "misperceptions correction" treatment revealing high community support for social distancing. The treatment effect on social distancing exhibits the spatial heterogeneity predicted by theory: negative in areas with low infection rates (reflecting the dominance of free-riding effects), and more positive in areas with higher rates (as perceived-infectiousness effects become increasingly prominent). In the area with the most cases, amounting to one-quarter of our sample, the treatment effect is positive and large in magnitude. The treatment effect on the expected future infection rate shows similar heterogeneity, confirming an additional theoretical prediction.

Our results suggest that when local infection rates are high, health policies shifting perceptions of community social distancing support upwards could help promote social distancing. These findings may also help predict the impacts of analogous public health messaging revealing community support for preventive measures against other infectious diseases.

References

Andersson, O., P. Campos-Mercade, A. Meier, and E. Wengstrom (2021). Anticipation of COVID-19 Vaccines Reduces Social Distancing. *SSRN Working Paper* (3765329).

Banerjee, A., E. La Ferrara, and V. H. Orozco-Olvera (2019). The Entertaining Way to Behavioral Change: Fighting HIV with MTV. *NBER Working Paper* (26096).

Benabou, R. and J. Tirole (2011). Identity, Morals, and Taboos: Beliefs as Assets. *Quarterly Journal of Economics 126*, 805–855.

Bicchieri, C. and E. Dimant (2019). Nudging with care: The risks and benefits of social information. *Public choice*, 1–22.

Bursztyn, L., A. L. Gonzalez, and D. Yanagizawa-Drott (2018). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review 110*, 2997–3029.

Cato, S., T. Iida, K. Ishida, A. Ito, K. McElwain, and M. Shoji (2020). Social Distancing as a Public Good Under the COVID-19 Pandemic. *Public Health 188*, 51–53.

Hershey, J. C., D. A. Asch, T. Thumasathit, J. Meszaros, and V. W. Waters (1994). The Roles of Altruism, Free Riding, and Bandwagoning in Vaccination Decisions. *Organizational Behavior and Human Decision Processes 59*, 177–187.

Jakubowski, A., D. Egger, C. Nekesa, L. Lowe, M. Walker, and E. Miguel (2021). Self-Reported Mask Wearing Greatly Exceeds Directly Observed Use: Urgent Need for Policy Intervention in Kenya. *medRxiy Preprint*.

Janzwood, S. (April 27, 2020). *The Social Distancing Norm Cascade: The Role of Belief Systems in Accelerating Normative Change during the COVID-19 Pandemic*. Canada: Cascade Institute.

Jones, S., E. Egger, and R. Santos (2020). Is Mozambique Prepared for a Lockdown During the COVID-19 Pandemic? *UNU-WIDER Blog*.

Lau, K., M. Miraldo, M. M. Galizzi, and K. Hauck (2019). Social Norms and Free-riding in Influenza Vaccine Decisions in the UK: An Online Experiment. *The Lancet 394*, 65.

Martinez, D., C. Parilli, C. Scartascini, and A. Simpser (2021). Let's (not) get together! The role of social norms on social distancing during COVID-19. *PLoS ONE 16*, 1–14.

Ministry of Health (2020). *COVID-19 Epidemiological Situation, Republic of Mozambique*. Maputo, Mozambique: Government of Mozambique.

National Institute of Statistics (INE) (2017). *Mozambique Population and Housing Census 2017*. Maputo, Mozambique: Government of Mozambique.

Nyusi, F. J. (August 5, 2020). *Communication to the Nation of His Excellency Philip Jacinto Nyusi, President of Republic of Mozambique, on the New State of Emergency, within the Scope of the Coronavirus Pandemic COVID-19*. Maputo, Mozambique: Maputo Mozambique.

Republic of Mozambique (April 2, 2020). *"Bulletin of the Republic", I Series, No. 64*. Maputo, Mozambique.

Schultz, W. P., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science 18*, 429–434.

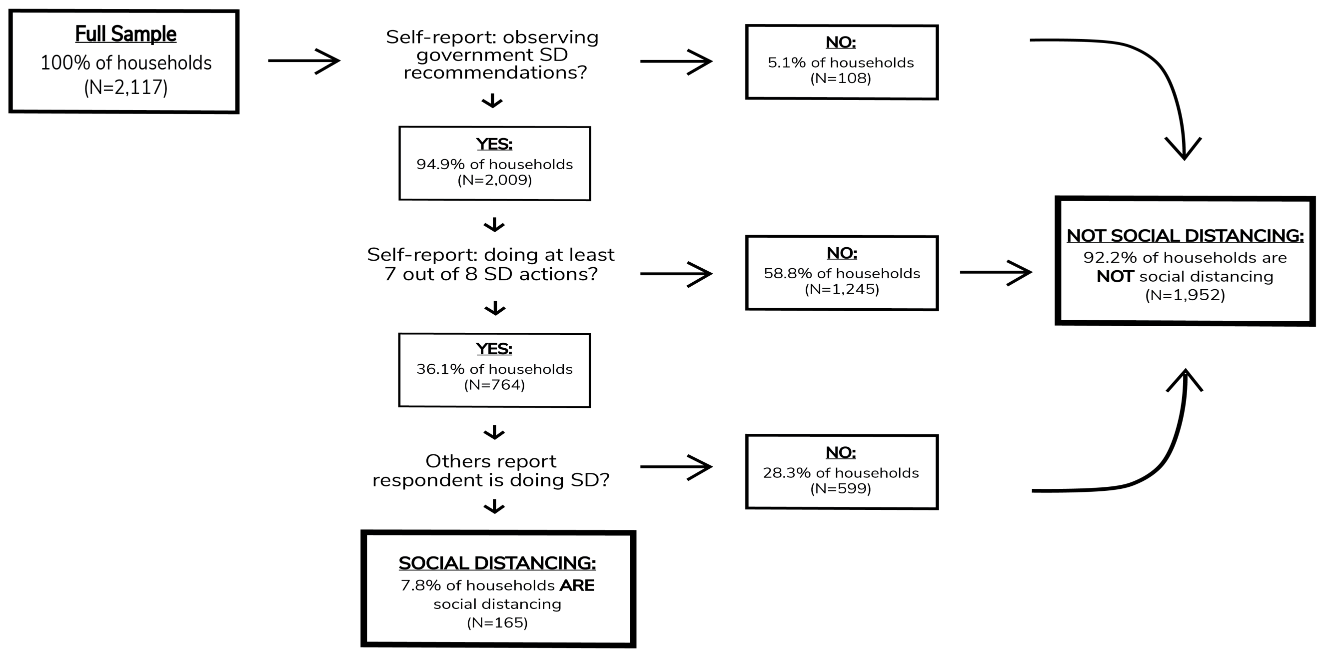
Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics 12*(1), 81–106.

U.S Embassy in Mozambique (2020). COVID-19 Information.

Yang, D., J. Allen IV, A. Mahumane, J. Riddell IV, and H. Yu (2021). Knowledge, Stigma and HIV Testing: An Analysis of a Widespread HIV/AIDS Program. *NBER Working Paper* (28716).

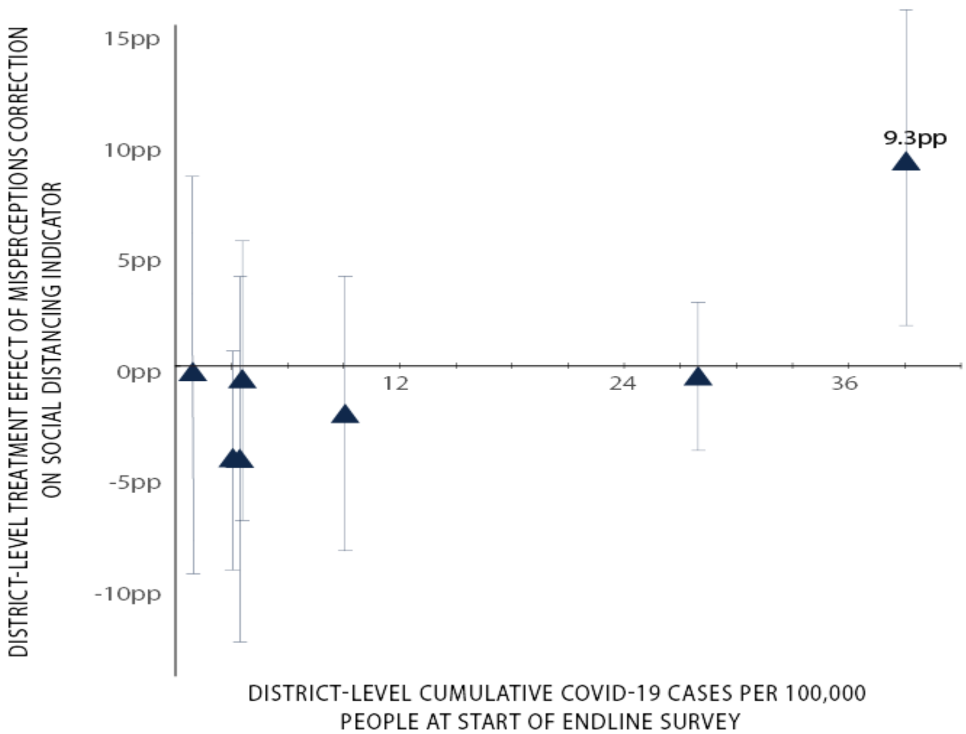
Yu, H. (2020). *Three Essays in Development Economics. Dissertation.* Ann Arbor, MI: University of Michigan.

Figure 1: **The Social Distancing Measure**



Notes: As pre-specified, respondents considered social distancing (SD) if: 1) self-report they are following government SD recommendations, 2) self-report they are doing at least seven out of eight SD actions, and 3) be reported by others in community to be SD. Percentages reported are all shares of full sample (N=2,117). See Table 1 and Section 3.2 and Appendix F of main text for social distancing question definitions.

Figure 2: **District-Level Misperceptions Correction Treatment Effects by Cumulative COVID-19 Cases**



Notes: Misperceptions correction treatment effects (triangles) estimated separately for each of seven districts (with 95% confidence intervals). District-level treatment effects plotted on vertical axis against district-level COVID-19 case loads at start of endline survey (per 100,000 population) on horizontal axis.

Table 1: **Summary Statistics of Social Distancing Support, Perceptions, and Behavior**

	VARIABLES	N	Mean	SD	Min	Max
(1)	Pre-baseline: Respondent supports social distancing (SD)	2,117	0.976	0.153	0	1
(2)	Pre-baseline: Perceived share of community supporting SD	2,109	0.689	0.313	0	1
(3)	Respondent supports SD	2,117	0.989	0.104	0	1
(4)	Perceived share of community supporting SD	2,114	0.800	0.262	0	1
(5)	Primary SD indicator	2,117	0.0784	0.269	0	1
(6)	Self-report of SD indicator	2,117	0.355	0.479	0	1
(7)	Self-report: Followed govt rules in past 14 days	2,117	0.949	0.219	0	1
(8)	Self-report: SD behaviors above median	2,117	0.361	0.481	0	1
(9)	Others' report of SD indicator	2,117	0.232	0.422	0	1
(10)	Other households' report of SD	2,117	0.378	0.485	0	1
(11)	Leaders' report of SD	2,117	0.519	0.500	0	1

Notes: Pre-baseline data collected from July 10 to August 16, 2020. All other summary statistics are baseline data collected from August 26 to October 4, 2020. Variables are as follows. Rows 1 & 3: indicator equal to one if respondent answers "yes" to supporting "the practice of social distancing to prevent the spread of coronavirus" and zero otherwise. Rows 2 & 4: perceived share of households (asked as "for every 10 households") in community that support social distancing (SD). Row 5: indicator for SD equal to one if respondent is SD according to self (Row 6) and others' reports (Row 9), and zero otherwise. Row 6: indicator for SD according to self if respondent answered "yes" to observing the government's recommendations on SD in the last 14 days (Row 7) and the report carrying out at least seven out of eight (above the sample median) social distancing behaviors (Row 8), and zero otherwise. Row 9: indicator for SD according to others if all other respondents (Row 10) and community leaders (Row 11) reported not knowing the respondent household, not seeing the respondent household in the past 14 days, or—if seen—that the respondent household 1) did NOT come closer than 1.5 meters to others outside their household; 2) did NOT shake hands, try to shake hands, or touch others outside their household; and 3) appeared to be observing the government's recommendations on SD, and zero otherwise.

Table 2: **Treatment Effects on Social Distancing and Expected COVID-19 Illnesses**

VARIABLES	(1) Primary SD Indicator	(2) Primary SD Indicator	(3) Perceived share of people in community that will get sick from Covid	(4) Perceived share of people in community that will get sick from Covid
T1: Misperceptions Correction	0.00425	-0.0466**	0.0418	-0.194**
	(0.0140)	(0.0191)	(0.0322)	(0.0944)
T2: Leader Endorsement	-0.00541	-0.0258	-0.0209	-0.0598
	(0.0137)	(0.0198)	(0.0308)	(0.0944)
T1 × District Covid Cases		0.00304***		0.00725**
		(0.00106)		(0.00291)
T2 × District Covid Cases		0.00122		0.00127
		(0.000992)		(0.00287)
Observations	2,117	2,117	812	812
R-squared	0.158	0.163	0.146	0.152
Control Mean DV	0.0857	0.0857	0.359	0.359
Control SD DV	0.280	0.280	0.369	0.369

Notes: Dependent variable in Columns 1-2 defined in Table 1. Dependent variable in Columns 3-4 is the expected future infection rate: "For every 10 people in your community, how many do you think would get sick from coronavirus?" (converted to share from 0 to 1). "T1: Misperceptions Correction" is equal to one if respondent was randomly assigned to the misperceptions correction treatment, and zero otherwise. "T2: Leader Endorsement" is equal to one if respondent was randomly assigned to the leader endorsement treatment, and zero otherwise. "T1 x District Covid Cases" & "T2 x District Covid Cases" are the respective treatment indicators interacted with district-level cumulative COVID-19 cases per 100,000 population at the start of the endline survey (as detailed in Appendix J). All regressions control for a baseline measure of the dependent variable, a vector of indicators for number of community leaders knowing the respondent at baseline (0 through 4), and a vector of indicators for number of other respondents knowing the respondent at baseline (0 through 8). All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Online Appendix

A Proofs

A.1 Proof of Theorem 1

The agent will adjust her effort level in response to the treatment to $\sqrt{\hat{e}} = \frac{\hat{\alpha}C}{2}H(\hat{x})$ where $H(x) = 1 - x(1 - \frac{1}{\sqrt{2}})$. Hence, the prior and posterior effort levels satisfy:

$$\frac{\sqrt{\tilde{e}}}{\sqrt{e}} = \frac{\tilde{\alpha}}{\alpha}\frac{H(\tilde{x})}{H(x)} \tag{A.1}$$

We take the ratios of Equations 6 and 7:

$$\frac{\hat{\alpha}}{\alpha} = \frac{1 - \sqrt{e}G(x)}{1 - \sqrt{e}G(\hat{x})} \tag{A.2}$$

We therefore obtain:

$$\frac{\sqrt{\tilde{e}}}{\sqrt{e}} = \frac{H(\hat{x})(1 - \sqrt{e}G(x))}{H(x)(1 - \sqrt{e}G(\hat{x}))} \tag{A.3}$$

Effort increases iff $\frac{\sqrt{\tilde{e}}}{\sqrt{e}} > 1$:

$$\begin{aligned} \frac{H(\hat{x})(1 - \sqrt{e}G(x))}{H(x)(1 - \sqrt{e}G(\hat{x}))} &> 1 \\ \sqrt{e}\left[H(x)G(\hat{x}) - H(\hat{x})G(x)\right] &> H(x) - H(\hat{x}) \end{aligned}$$

Now note that $G(x) = 2xH(x)$ such that:

$$\begin{aligned} \sqrt{e}2H(x)H(\hat{x})(\hat{x} - x) &> (1 - \frac{1}{\sqrt{2}})(\hat{x} - x) \qquad \text{(A.4)} \\ \sqrt{e} &> \frac{1 - \frac{1}{\sqrt{2}}}{H(x)H(\hat{x})} \end{aligned}$$

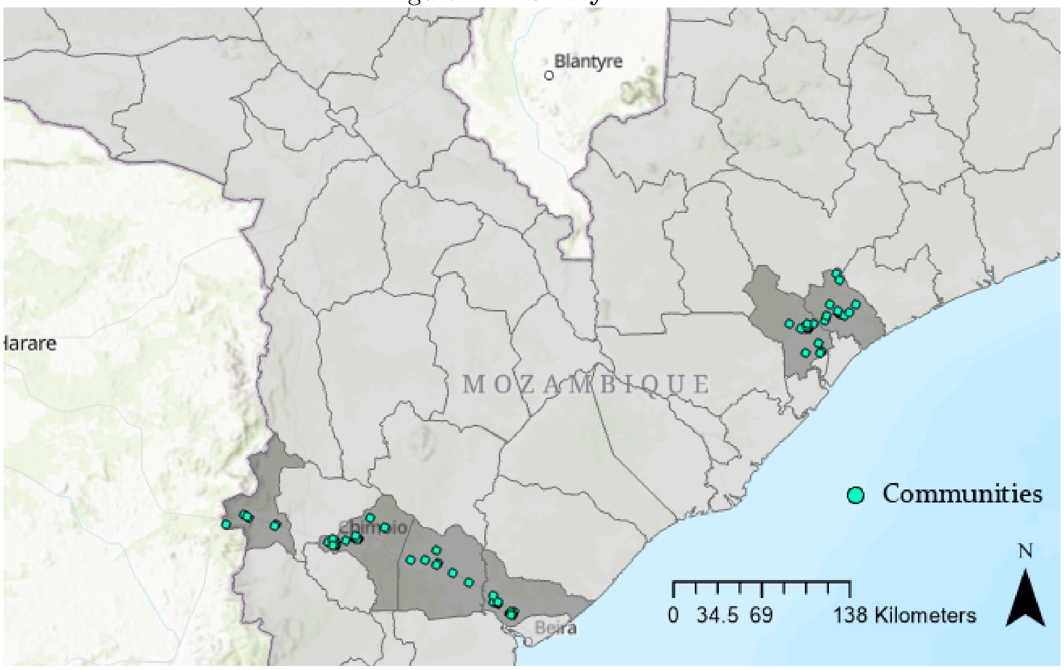
This shows that the perceived-infectiousness effect dominates if the initial effort level e is high enough. Effort is determined by Equation 5 and increases with α (which increases with $\hat{\alpha}$). Therefore, for sufficiently large $\hat{\alpha}$ the perceived-infectiousness effect dominates.

B Study Area and Timeline

The Mozambican government declared a State of Emergency due to the COVID-19 pandemic on March 31, 2020, recommending social distancing (at least 1.5 meters) and requiring it at public and private institutions and gatherings. The government also suspended schools, required masks at funerals and markets, banned gatherings of 20 or more, and closed bars, cinemas and gymnasiums (Republic of Mozambique, 4/1/2020). The government stopped short of implementing a full economic "lockdown" due to its economic costs (Jones et al., 2020). On August 5, 2020, the government renewed the State of Emergency, called for improved mask-wearing, and announced a schedule for loosening restrictions (Nyusi, 8/5/2020). In September, the government loosened some restrictions including resuming religious gatherings at 50% capacity (U.S Embassy in Mozambique). Throughout this period, the government's social distancing recommendation remained constant.

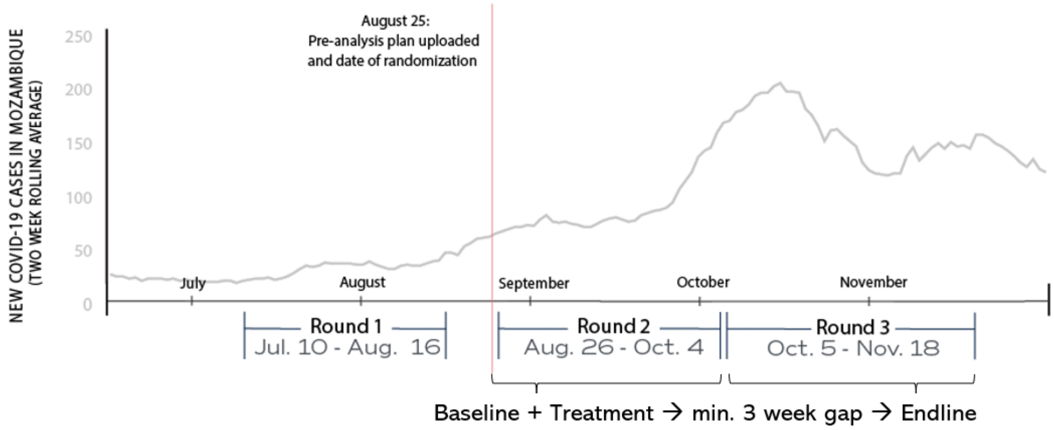
Study participants come from 76 communities in central Mozambique. The study communities are in seven districts of three provinces: Dondo and Nhamatanda in Sofala province; Gondola, Chimoio and Manica in Manica province; and Namacurra and Nicoadala in Zambezia province. These 76 communities are mapped in Figure A.1. Compared to other communities in Mozambique, the study areas are relatively accessible to main transport corridors (highways and ports), and are thus important geographic conduits for infectious disease.

Figure A.1: **Study Area**



We collected survey data in three rounds between July 10 and November 18, 2020. Appendix Figure A.2 depicts the study timeline below a rolling average of new Mozambican COVID-19 cases. We piloted surveys in Round 1. Immediately before the Round 2 survey, we randomly assigned households to treatments and submitted our pre-analysis plan to the AEA RCT Registry. The Round 2 survey served as a baseline, and was immediately followed (on the same phone call) by our treatment interventions. Round 3 was our endline survey. Surveys collected data on COVID-19 knowledge, beliefs, and behaviors. While data collection for Round 3 began only one day after completion of Round 2, there was a minimum of 3.0 weeks and average of 6.3 weeks between Rounds 2 and 3 surveys for any given respondent. While the Round 1 survey occurred when new COVID-19 cases remained relatively steady, both the Round 2 and Round 3 surveys occurred during a period of substantial growth in new COVID-19 cases.

Figure A.2: **Study Timeline**



Notes: Round 1 is pre-baseline survey to collect social distancing support data, Round 2 is baseline survey, and Round 3 is endline survey. There is at least a three week gap between baseline and endline survey for any given study participant. Pre-analysis plan uploaded and treatments randomly assigned immediately prior to start of Round 2 baseline survey, on Aug. 25, 2021. Treatments implemented immediately following baseline survey on same phone call. Baseline measures reported in Table 1 come from Round 2 surveys and endline measures come from Round 3 surveys.

C Treatment Details and Scripts

Both the misperceptions correction and leader endorsement treatments were implemented directly following the baseline survey, on the same phone call. If a respondent was randomly assigned to a treatment, the corresponding intervention text would appear on the enumerator's tablet. Enumerators read a script aloud exactly as shown below. Following the treatment, respondents were asked if they would like the information repeated. Of the N=628 receiving the misperceptions correction and N=637 receiving the leader endorsement, only 8.6% and 9.4% asked for the script to be repeated, respectively.

Misperceptions correction treatment: The misperceptions correction treatment involves sharing the level of "actual community support for social distancing" with the respondent if they underestimate or correctly estimate that level. We express this as integer units out of 10, rounded to the nearest integer based on actual shares of respondents in the community expressing support for social distancing in the Round 1 (pre-baseline) survey. In 63 out of 76 communities (82.9%) the number we convey to respondents is 10 out of 10, and in 13 communities (17.1%) the number is 9 out of 10.

Script for T1: Misperceptions Correction – "Now I want to give you some information about social distancing. In this survey, you indicated that you think *<insert respondent's answer here>* of every 10 households in your community support the practice of social distancing."

- *If response UNDERESTIMATES community support for social distancing:* "However, more households support social distancing than you think! Based on the results of our first COVID-19 survey, approximately *<insert actual community support for social distancing here>* of every 10 households in your community support social distancing to prevent the spread of the coronavirus."

- *If response CORRECTLY ESTIMATES community support for social distancing:* "You are correct! Based on the results of our first COVID-19 survey, approximately *<insert actual community support for social distancing here>* of every 10 household in your community support social distancing to prevent the spread of the coronavirus."

- *If response OVERESTIMATES community support for social distancing: (no information given)*

Leader endorsement treatment: We identified and surveyed community opinion leaders prior to the baseline survey and requested their permission to tell others in their community that they "support social distancing, are practicing social distancing, and encourage others to do the same". Then, in this treatment, we reported this endorsement to respondents, mentioning the community leader(s) by name.

Script for T2: Leader Endorsement – "Our research team recently called and talked to your *<list leaders' titles and names here>*. They said that they support social distancing, are practicing social distancing themselves, and encourage others to do the same."

D Populated Pre-analysis Plan

On August 25, 2020, prior to baseline data collection, we uploaded our pre-analysis plan (PAP) to the American Economic Association's RCT Registry, registration ID number AEARCTR-0005862: https://doi.org/10.1257/rct.5862-1.0.

In our PAP, we specify the following regression for our primary analysis, which is the same as Equation 8 in the main text:

$$Y_{ijd} = \beta_0 + \beta_1 T1_{ijd} + \beta_2 T2_{ijd} + \eta B_{ijd} + \delta_{ijd}^{others} + \delta_{ijd}^{leaders} + \gamma_{jd} + \varepsilon_{ijd} \tag{D.1}$$

where Y_{ijd} is the social distancing indicator for household i in community j and district d; $T1_{ijd}$ and $T2_{ijd}$ are indicator variables for the misperceptions correction and leader endorsement treatment groups, respectively; B_{ijd} is the baseline value of the dependent variable; γ_{jd} are community fixed effects; and ε_{ijd} is a mean-zero error term. We report robust standard errors. The regression also controls for the number of other survey respondents and community leaders who report knowing the survey respondent at baseline (in Round 2). Specifically, δ_i^{others} is a vector of dummy variables for the distinct number of other surveyed study respondents who report knowing the household (0, 1, 2. . . , 7, 8 or more; where 8 is the first integer where over 90% of the sample is represented by previous non-negative integers), and $\delta_i^{leaders}$ is a vector of dummy variables for the distinct number of community leaders who report knowing the household (0, 1, 2, 3, 4; where 4 is maximum number of leaders found within one of the 76 sample communities). Including this control variable helps reduce residual variance in the dependent variable, because respondents who are known by more others in the community will also have more reports of social interactions with others. These results are presented in the main paper in Table 2 column (1) and are also replicated in column (1) of Table A.1.

Additionally, we pre-specified the following secondary analyses. First, we analyze impacts of the social distancing treatments on the separate components of the social distancing index—the others' and self-report. These results are presented in Table A.1 columns (2) & (3), respectively. Treatment effects on these outcomes are very similar to those in column (1).

Second, we also pool SD1 and SD2 together to examine the effect of some endorsement of social distancing (whether by other community members or by community leaders) on the primary social distancing outcome. These coefficient in Table A.1 column (4) is small in magnitude and not statistically significantly different from zero at conventional levels.

Table A.1: **Additional Prespecified Analyses**

VARIABLES	(1) Primary SD Indicator	(2) Others' Report of SD	(3) Self-Report of SD	(4) Primary SD Indicator
T1: Misperceptions Correction	0.00425	0.000950	0.0134	
	(0.0140)	(0.0181)	(0.0238)	
T2: Leader Endorsement	-0.00541	0.0145	-0.0189	
	(0.0137)	(0.0183)	(0.0234)	
Pooled SD Treatments				-0.000642
				(0.0116)
Observations	2,117	2,117	2,117	2,117
R-squared	0.158	0.333	0.211	0.158
Control Mean DV	0.0857	0.211	0.406	0.0857
Control SD DV	0.280	0.408	0.491	0.280

Notes: Dependent variables are defined in Table 1. "T1: Misperceptions Correction" is an indicator equal to one if respondent was randomly assigned to the misperceptions correction treatment, and zero otherwise. "T2: Leader Endorsement" is an indicator equal to one if respondent was randomly assigned to the leader endorsement treatment, and zero otherwise. "Pooled SD Treatments" is an indicator equal to one if respondent was randomly assigned to the misperceptions correction treatment or leader endorsement treatment, and zero otherwise. Controls are as defined in Table 2. All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

We also randomly assigned a family of treatments to improve COVID-19 knowledge in the same study population.[1] Randomization of the misperceptions correction and leader endorsement treatments were stratified within 76 communities and within the separate knowledge treatment conditions (i.e., the knowledge and social distancing treatments were cross-randomized). As pre-specified, we run a regression on the primary social distancing outcome with indicators for social distancing treatments, the cross-randomized knowledge treatments and their interaction terms. Results are presented in Table A.2, and show no large or statistically significant interaction effects between the social distancing and knowledge treatments.

[1]The pre-analysis plan (PAP) for the knowledge study can be found here: https://fordschool.umich.edu/mozambique-research/combatting-covid-19.

Table A.2: **Interactions between Social Distancing and Knowledge Treatments**

	(1)
VARIABLES	Primary SD Indicator
T1: Misperceptions Correction	-0.0237
	(0.0214)
T2: Leader Endorsement	-0.0210
	(0.0222)
K1: Incentive	-0.0218
	(0.0241)
K2: Feedback	-0.00250
	(0.0251)
K3: Incentive & Feedback	-0.0144
	(0.0238)
T1 × K1	0.0545
	(0.0390)
T2 × K1	0.0249
	(0.0372)
T1 × K2	0.0467
	(0.0397)
T2 × K2	0.0139
	(0.0385)
T1 × K3	0.0404
	(0.0382)
T2 × K3	0.0374
	(0.0372)
Observations	2,117
R-squared	0.160
Control Mean DV	0.0857
Control SD DV	0.280

Notes: Dependent variable is defined in Table 1. Social distancing treatments are defined in Table 2. "K1 Incentive", "K2 Feedback", and "K3 Incentive & Feedback" are indicators equal to one if respondent was randomly assigned to one of these knowledge treatments, and zero otherwise. Remaining regressors represent interactions between social distancing treatments and the knowledge treatments. Controls are as defined in Table 2. Regression also includes community fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

E Attrition and Balance

Appendix Table A.3 presents regressions examining whether attrition and baseline variables are balanced with respect to treatment assignment[2]. Attrition between Round 2 (baseline) and Round 3 (endline) is only 4.9% and is less than 5.6% in each of the seven districts surveyed. Balance in attrition is confirmed in column 1, which starts with the Round 2 (baseline) sample and regresses treatments on an indicator equal to one if the respondent was not reached for the Round 3 (endline) survey. Balance in baseline social distancing outcomes is confirmed in columns 2-4, which examines the Round 2 social distancing outcomes. Balance in baseline household characteristics is confirmed in columns 6-8, which examines the final Round 3 sample and regresses treatments on Round 1 measures of household income, an index of food insecurity, and an indicator for presence of an older adult over 60 years. In not a single regression in the table is a coefficient on a treatment indicator statistically significant at conventional levels.

[2]Figure A.2 shows the study timeline for the three survey rounds collected. Round 1 is a pre-baseline measure, Round 2 measures baseline values and Round 3 measures endline outcomes.

Table A.3: **Treatment Effect on Attrition and Balance**

VARIABLES	(1) Attrition	(2) Primary SD Indicator	(3) Others' Report of SD	(4) Self-Report of SD	(5) Perceived Social Norm	(6) Hh Income	(7) Food Insecurity	(8) Older Adult in Hh
T1: Misperceptions Correction	-0.0127	-0.0176	-0.000450	-0.00956	-0.0101	-159.5	0.00107	-0.00293
	(0.0111)	(0.0134)	(0.0203)	(0.0247)	(0.0138)	(181.7)	(0.0191)	(0.0250)
T2: Leader Endorsement	-0.00154	-0.00324	0.00897	0.00420	-0.0201	-39.95	-0.0240	0.0240
	(0.0113)	(0.0143)	(0.0206)	(0.0249)	(0.0137)	(181.8)	(0.0193)	(0.0252)
Observations	2,226	2,117	2,117	2,117	2,114	1,873	2,117	2,096
R-squared	0.030	0.096	0.199	0.076	0.047	0.043	0.090	0.058
Control Mean DV	0.0533	0.0833	0.229	0.356	0.810	1176	0.842	0.342
Control SD DV	0.225	0.277	0.420	0.479	0.262	4029	0.365	0.475

Notes: Dependent variables are as follows. Column 1: indicator if respondent attrited from the sample between baseline and endline. Columns 2-4: baseline SD outcomes defined in Table 1. Column 5: baseline perceived share of community supporting SD, defined further in Table 1. Column 6: at pre-baseline, self-reported total income for the previous week (in Mozambican meticais). Column 7: indicator if, in the last 7 days, household has 1) lacked food; 2) reduced number of meals/portions; or was unable to buy their usual amount of food due to 3) market shortages, 4) high prices, 5) reduced income. Column 8: indicator if adult age 60 or older is present in the household. Controls are as defined in Table 2. All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

F Social Distancing Index

The list of actions included in the Social Distancing Index and their corresponding summary statistics are presented below.

Social Distancing Actions: Is this something your household has been doing for the last seven days? (Answers indicating social distancing in parentheses.[3])

1. Shop in crowded areas like informal markets (No)
2. Gather with several friends (No)
3. Help the elderly avoid close contact with other people, including children (Yes)
4. If show symptoms of coronavirus, immediately inform my household and avoid people (Yes)
5. Drink alcohol in bars (No)
6. Wear a face mask if showing symptoms of coronavirus (Yes)
7. Instead of meeting in person, call on the phone or send text message (Yes)
8. Allow children to build immunity by playing with children from other households (No)

Below are the summary statistics for the questions that comprise the self-reported social distancing index at baseline and endline. Respondents were asked "Is this something your household has been doing for the last seven days?" about a randomly determined four social distancing actions at baseline and all eight social distancing actions at endline. Responses were coded as indicators equal to one if indicative of social distancing (answers that indicate social distancing shown in parentheses), and zero otherwise.

[3]For items 4 and 6 that are conditional on showing symptoms, survey staff instructed respondents to answer "Yes" (doing social distancing) if not showing symptoms.

Table A.4: **Summary Statistics for Components of Social Distancing Index**

	Baseline			Endline		
VARIABLES	N	Mean	SD	N	Mean	SD
Shop in crowded areas like informal markets (No)	1,032	.642	.479	2,115	.678	.467
Gather with several friends (No)	1,047	.349	.477	2,113	.414	.493
Help the elderly avoid close contact with other people, including children (Yes)	1,094	.877	.329	2,114	.923	.266
If show symptoms of coronavirus, immediately inform my household and avoid people (Yes)	1,050	.836	.370	2,113	.859	.347
Drink alcohol in bars (No)	1,082	.226	.419	2,113	.272	.445
Wear a face mask if showing symptoms of coronavirus (Yes)	1,034	.902	.297	2,114	.885	.319
Instead of meeting in person, call on the phone or send text message (Yes)	1,039	.935	.247	2,112	.930	.255
Allow children to build immunity by playing with children from other households (No)	1,070	.439	.497	2,113	.456	.498

Notes: Variables are coded as indicators equal to one if indicative of social distancing (answers that indicate social distancing shown in parentheses), and zero otherwise. Respondents were asked "Is this something your household has been doing for the last seven days?" about a randomly determined four social distancing actions at baseline and all eight social distancing actions at endline. The baseline sample was asked a subset of these questions which explains the smaller number of observations at baseline.

G Treatment Effect Estimates from Logit and Probit Regressions

The primary social distancing indicator is a binary variable that is analyzed using an ordinary least-squares (OLS) regression, as prespecified. As a robustness check, we adapt Equation 8 to be run using logit and probit regression.

Table A.5 presents results from the logistic regression on the primary outcomes, while Table A.6 presents corresponding probit regression results. Regression coefficients are presented as marginal effects. Results in both tables are consistent with the results from OLS linear probability models presented in Table 2.

Table A.5: **Treatment Effects Estimated Using Logistic Regression**

VARIABLES	(1) Primary SD Indicator	(2) Primary SD Indicator	(3) Perceived share of households in community that will get sick from Covid	(4) Perceived share of households in community that will get sick from Covid
T1: Misperceptions Correction	0.0100	-0.0756**	0.0270	-0.403***
	(0.0221)	(0.0376)	(0.0395)	(0.138)
T2: Leader Endorsement	-0.00695	-0.0398	-0.0274	-0.300**
	(0.0222)	(0.0349)	(0.0394)	(0.135)
T1 × District Covid Cases		0.00384***		0.0132***
		(0.00131)		(0.00397)
T2 × District Covid Cases		0.00162		0.00842**
		(0.00129)		(0.00397)
Observations	1,285	1,285	806	806
Control Mean DV	0.141	0.141	0.356	0.356
Control SD DV	0.349	0.349	0.368	0.368

Notes: Dependent variables are defined in Tables 1 and 2. Coefficients presented are marginal effects from logit regression. Social distancing treatments and controls are as defined in Table 2. All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: **Treatment Effects Estimated Using Probit Regression**

VARIABLES	(1) Primary SD Indicator	(2) Primary SD Indicator	(3) Perceived share of households in community that will get sick from Covid	(4) Perceived share of households in community that will get sick from Covid
T1: Misperceptions Correction	0.00892	-0.0709**	0.0288	-0.401***
	(0.0212)	(0.0347)	(0.0390)	(0.132)
T2: Leader Endorsement	-0.00837	-0.0356	-0.0298	-0.306**
	(0.0214)	(0.0330)	(0.0392)	(0.135)
T1 × District Covid Cases		0.00369***		0.0132***
		(0.00126)		(0.00384)
T2 × District Covid Cases		0.00139		0.00851**
		(0.00124)		(0.00396)
Observations	1,285	1,285	806	806
Control Mean DV	0.141	0.141	0.356	0.356
Control SD DV	0.349	0.349	0.368	0.368

Notes: Dependent variables are defined in Tables 1 and 2. Coefficients presented are marginal effects from probit regression. Social distancing treatments and controls are as defined in Table 2. All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

H Effect on Perceived Community Support

Using answers to the question "For every 10 households in your community, how many support social distancing?", we estimate each respondent's perceived share of households in the community who support social distancing. We note that this measure may be an upward-based estimate of true perceptions of community support, since experimenter demand effects may lead respondents to report higher shares of support for social distancing in order to make their communities look favorable.

Table A.7 presents the cumulative distribution of this perceived community support measure in the full samples at baseline and endline, and subdivided by treatment arm at endline. Even at baseline, the distribution is skewed upwards with over 90% of the sample reporting that the majority (50% or greater) of households in their community support social distancing and over half of the sample reporting that 100% of households do the same. Therefore, this measure may be limited in its ability to measure the effect of a misperceptions correction given that there is "little room to improve" for many respondents in the sample. If some high estimates are due to experimenter demand effects, then the misperceptions correction may still boost respondents' true perception of community support in a way not captured by our measure.

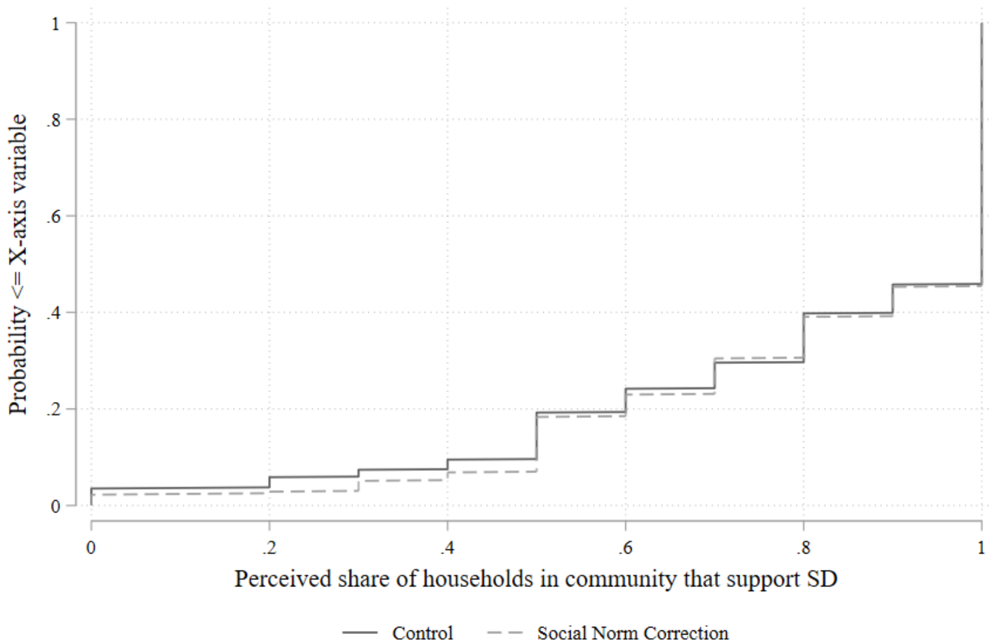
Table A.7: **Sample Distribution (Cumulative %) by Perceived Community Support**

	Baseline	Endline			
Perceived Share	Total	Total	Control	T1	T2
0%	2.7	2.8	3.5	2.2	2.5
10%	3.1	3.1	3.6	2.4	3.0
20%	4.4	4.4	5.9	2.9	3.9
30%	6.5	6.5	7.4	5.1	6.6
40%	9.6	8.8	9.5	6.9	9.9
50%	21.1	19.0	19.3	18.3	19.2
60%	27.1	23.9	24.2	23.0	24.5
70%	33.4	30.3	29.6	30.5	31.1
80%	43.4	40.8	39.8	39.1	44.0
90%	48.9	46.8	45.8	45.3	49.6
100%	100.0	100.0	100.0	100.0	100.0

Notes: Perceived share of households in the community who support social distancing is estimated by dividing responses to the question "For every 10 households in your community, how many support social distancing?" by 10, and hence has 11 categories from 0%, 10%... 90%, 100%. Cells report cumulative percentages from 0% up to the row in question. At baseline, "Total" refers to the Round 2 responses from the whole sample. At endline, "Total" refers to Round 3 responses from the whole sample, "Control" from the control group, "T1" from the misperceptions correction treatment group, and "T2" from the leader endorsement treatment group.

We find that the misperceptions correction treatment did increase respondents' perceived community support, particularly for those at the lower end of the distribution. Figure A.3 shows the cumulative distribution function for the perceived community support measure at endline. Relative to the control group, those receiving the misperceptions correction treatment were less likely to report that fewer than 50% of households in their community supported social distancing, instead reporting higher perceptions of community support. Further, Table A.8 presents three regressions estimating the treatment effects on the perceived community support. In column (1), the dependent variable is the perceived share of households in the community who support social distancing. The coefficient is positive and marginally statistically significant (p-value 0.12). Regressions in columns (2) and (3) find that the misperceptions correction treatment has a positive effect on an indicator for the respondent believing the majority (50% or more) of households in their community support social distancing, and an indicator that the respondent's perceived community support increased between baseline and endline (both coefficients are statistically significantly different from zero at the 5% level).

Figure A.3: **Cumulative Distribution of Perceived Community Support by Treatment**



Notes: Perceived share of households in the community who support social distancing is estimated by dividing responses to the question “For every 10 households in your community, how many support social distancing?” by 10, and hence has 11 categories from 0%, 10%... 90%, 100%. Figure depicts the cumulative distribution function of this variable for the “Control” group and “Misperceptions Correction” treatment arm. The leader endorsement treatment is excluded for clarity.

Table A.8: **Treatment Effects on Perceived Community Support (PCS)**

VARIABLES	(1) Continuous PSN	(2) Indicator if PSN $\geq$ 50%	(3) Indicator if PSN increased
T1: Misperceptions Correction	0.0196	0.0291**	0.0507**
	(0.0128)	(0.0138)	(0.0241)
T2: Leader Endorsement	0.00405	0.00357	0.0358
	(0.0128)	(0.0149)	(0.0236)
Observations	2,116	2,116	2,113
R-squared	0.164	0.118	0.043
Control Mean DV	0.812	0.905	0.255
Control SD DV	0.268	0.293	0.436

Notes: Dependent variables are defined as follows. Column 1 is the perceived share of households in community that support social distancing, which takes on the values shown in Table A.7. Column 2 is an indicator equal to one if respondent reports that majority (50% or more) of households in community support social distancing, and zero otherwise. Column 3 is an indicator equal to one if the respondent's perceived community support increased between the baseline (pre-treatment) and endline (post-treatment) surveys. "T1: Misperceptions Correction" & "T2: Leader Endorsement" and controls are as defined in Table 2, except column 3 does not include a baseline value of the outcome as a control as it was used to calculate the outcome. All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

I Excluding Chimoio District

A central finding of the paper is the heterogeneity in the treatment effect of the misperceptions correction treatment with respect to local COVID-19 cases per 100,000 population (Column 2 of Table 2). A question that arises is whether this heterogeneity is entirely driven by the Chimoio district, which has the highest case loads in the sample by a fair margin (see Figure 2 and Appendix J). We therefore test the robustness of our findings to excluding from the sample the 524 respondents in Chimoio district (one-quarter of the sample), thereby only exploiting the more limited variation in district-level case loads across the remaining six districts.

Table A.9 below presents coefficient estimates from estimating Equations 8 and 9 in this restricted sample. First of all, Column 1 reveals that the coefficient on the misperceptions correction treatment is negative and statistically significant at the 10% level. Because this sample drops the district with the highest case loads, this result is consistent with theoretical predictions and previous findings that at lower case loads, the misperceptions correction treatment effect is more likely to be negative.

In Column 2, where we test for heterogeneity in the treatment effect, results are quite similar to the findings in Column 2 of Table 2 in the main text. The T1 main effect and interaction term coefficients are of similar magnitudes to those in Column 2 of Table 2, and maintain statistical significance at conventional levels (the T1 interaction term coefficient is now significant at the 10% instead of 5% level).

In sum, our central findings regarding heterogeneity in the treatment effect of the misperceptions correction treatment are robust to excluding from the sample respondents from the district (Chimoio) with the highest COVID-19 case loads.

Table A.9: **Treatment Effects on Social Distancing, Excluding Chimoio District**

	(1)	(2)
VARIABLES	Primary SD Indicator	Primary SD Indicator
T1: Misperceptions Correction	-0.0237*	-0.0410**
	(0.0131)	(0.0194)
T2: Leader Endorsement	-0.0150	-0.0263
	(0.0141)	(0.0208)
T1 $\times$ District Covid Cases		0.00186*
		(0.000989)
T2 $\times$ District Covid Cases		0.00120
		(0.00103)
Observations	1,593	1,593
R-squared	0.141	0.142
Control Mean DV	0.0710	0.0710
Control SD DV	0.257	0.257

Notes: Regressions exclude 524 respondents from Chimoio district. Dependent variable in Columns 1-2 defined in Table 1. "T1: Misperceptions Correction" is equal to one if respondent was randomly assigned to the misperceptions correction treatment, and zero otherwise. "T2: Leader Endorsement" is equal to one if respondent was randomly assigned to the leader endorsement treatment, and zero otherwise. "T1 x District Covid Cases" & "T2 x District Covid Cases" are the respective treatment indicators interacted with district-level cumulative COVID-19 cases per 100,000 population at the start of the endline survey (as detailed in Appendix J). All regressions control for a baseline measure of the dependent variable, a vector of indicators for number of community leaders knowing the respondent at baseline (0 through 4), and a vector of indicators for number of other respondents knowing the respondent at baseline (0 through 8). All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

J COVID-19 Cases

Data on district-level population come from Mozambique's 2017 Census (National Institute of Statistics (INE), 2017). District COVID-19 case counts come from the government's COVID-19 Mozambique dashboard (Ministry of Health, 2020) and correspondence with provincial health offices. Each district's case count is as of the start date of the endline survey in the district (ranging from October 5 to November 1, 2020). We also show the number of respondents in our study sample in each district.

Table A.10: **Covid Cases by District**

	(1)	(2)	(3)	(4)
	Cumulative Covid	Cases per	Population	Number of Study
DISTRICT	Cases	100,000 people		Respondents
Sofala Province				
Dondo	8	4.136	193,382	323
Nhamatanda	12	4.299	279,081	214
Manica Province				
Gondola	3	3.553	84,429	224
Chimoio	142	39.082	363,336	524
Manica	20	9.292	215,239	290
Zambezia Province				
Namacurra	4	1.652	242,126	244
Nicoadala	52	28.779	180,686	298

K Robustness of SD Measure

Table A.11: **Treatment Effects on Social Distancing and Expected COVID-19 Illnesses for Modified Social Distancing Measures**

VARIABLES	(1) Alternative SD Indicator 1	(2) Alternative SD Indicator 1	(3) Alternative SD Indicator 2	(4) Alternative SD Indicator 2
T1: Misperceptions Correction	-0.00596	-0.0459**	-0.00133	-0.0452**
	(0.0159)	(0.0215)	(0.0144)	(0.0198)
T2: Leader Endorsement	-0.00729	-0.0313	-0.00179	-0.0372*
	(0.0159)	(0.0226)	(0.0144)	(0.0204)
T1 × District Covid Cases		0.00238**		0.00262**
		(0.00120)		(0.00109)
T2 × District Covid Cases		0.00144		0.00213**
		(0.00117)		(0.00106)
Observations	2,117	2,117	2,117	2,117
R-squared	0.197	0.199	0.163	0.167
Control Mean DV	0.124	0.124	0.0939	0.0939
Control SD DV	0.330	0.330	0.292	0.292

Notes: Dependent variable in Columns 1-2 is social distancing measure in which threshold number of self-reported actions to be considered social distancing is 6 out of 8 (instead of 7 out of 8). Dependent variable in Columns 3-4 is social distancing measure which excludes social distancing actions 4 and 6 that are conditional on experiencing symptoms (threshold number of self-reported actions is changed to 5 out of 6 actions). "T1: Misperceptions Correction" is equal to one if respondent was randomly assigned to the misperceptions correction treatment, and zero otherwise. "T2: Leader Endorsement" is equal to one if respondent was randomly assigned to the leader endorsement treatment, and zero otherwise. "T1 x District Covid Cases" & "T2 x District Covid Cases" are the respective treatment indicators interacted with district-level cumulative COVID-19 cases per 100,000 population at the start of the endline survey (as detailed in Appendix J). All regressions control for a baseline measure of the dependent variable, a vector of indicators for number of community leaders knowing the respondent at baseline (0 through 4), and a vector of indicators for number of other respondents knowing the respondent at baseline (0 through 8). All regressions also include community fixed effects. Robust standard errors in parentheses. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

L Clustering Standard Errors

Table A.12: **Treatment Effects on Social Distancing and Expected COVID-19 Illnesses, with Clustered Standard Errors**

	Primary SD Indicator			
	(1)	(2)	(3)	(4)
VARIABLES	Clustered at Community	Clustered at District	Clustered at Community	Clustered at District
T1: Misperceptions Correction	0.00425	0.00425	-0.0466***	-0.0466***
	(0.0133)	(0.0269)	(0.0150)	(0.0124)
T2: Leader Endorsement	-0.00541	-0.00541	-0.0258	-0.0258
	(0.0132)	(0.0133)	(0.0169)	(0.0163)
T1 × District Covid Cases			0.00304***	0.00304***
			(0.000911)	(0.000514)
T2 × District Covid Cases			0.00122	0.00122**
			(0.00103)	(0.000462)
Observations	2,117	2,117	2,117	2,117
R-squared	0.158	0.158	0.163	0.163
Control Mean DV	0.0857	0.0857	0.0857	0.0857
Control SD DV	0.280	0.280	0.280	0.280

Notes: Standard errors (in parentheses) are clustered at the level of 76 communities (Columns 1 and 3) or 7 districts (Columns 2 and 4). Dependent variable is defined in Table 1. "T1: Misperceptions Correction" is equal to one if respondent was randomly assigned to the misperceptions correction treatment, and zero otherwise. "T2: Leader Endorsement" is equal to one if respondent was randomly assigned to the leader endorsement treatment, and zero otherwise. "T1 x District Covid Cases" & "T2 x District Covid Cases" are the respective treatment indicators interacted with district-level cumulative COVID-19 cases per 100,000 population at the start of the endline survey (as detailed in Appendix J). All regressions control for a baseline measure of the dependent variable, a vector of indicators for number of community leaders knowing the respondent at baseline (0 through 4), and a vector of indicators for number of other respondents knowing the respondent at baseline (0 through 8). All regressions also include community fixed effects. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.